

Завдання.

1. Для даних лаб. роботи №6 провести відбір пояснювальних ознак. При цьому припустити, що перших шість ознак $X_1 - X_6$ є пояснювальними, а остання – X_7 – результативною. Для цього вибірково розглянути
 - а) всі одноелементні підмножини (шість підмножин);
 - б) довільних дві двоелементних, триелементних, чотириелементних та п'ятиелементних підмножини, враховуючи розбиття вихідної множини ознак на однорідні підмножини (результати лаб. роботи №4);
 - в) шестиелементну підмножину.
2. Для кожної з підмножин обчислити значення N_m .
3. Серед відібраних підмножин обрати дві найкращих та побудувати регресійну залежність для кожної відібраної підмножини та перевірити по сумі квадратів залишкових різниць, яка з них є найкращою.
4. Зробити висновки.

Вхідні дані							
№ п/п	Номери ознак						
	1	2	3	4	5	6	7
1	1,85	36,72	40,5	45,15	490	22,36	185,25
2	1,95	35,53	54,54	55,47	580	24,96	191,1
3	1,92	38,42	64,8	72,24	710	27,04	253,5
4	1,9	37,4	67,5	86,43	740	38,22	308,1
5	0,35	38,08	3,78	47,73	450	4,68	74,1
6	0,35	7,65	4,32	55,47	500	5,98	95,55
7	0,55	6,97	6,48	76,11	680	9,36	95,55
8	0,57	10,88	6,48	87,72	760	10,92	120,9
9	0,86	12,41	28,62	33,54	480	16,12	136,5
10	0,86	12,75	29,7	38,7	590	32,5	163,8
11	0,84	12,41	34,02	42,57	600	54,34	154,05
12	0,82	11,73	37,26	51,6	640	106,86	154,05
13	0,82	16,49	8,64	64,5	340	12,74	169,65
14	0,82	15,13	9,18	78,69	410	14,82	216,45
15	0,81	15,3	10,26	94,17	450	16,64	212,55
16	0,8	14,96	11,88	104,49	500	60,84	366,6
17	1,19	19,04	32,94	45,15	580	18,46	136,5
18	1,2	20,23	34,56	55,47	660	19,76	169,65
19	1,29	22,1	43,2	68,37	780	23,66	222,3
20	1,27	21,25	45,9	77,4	830	26	200,85

Методичні вказівки до виконання роботи.

Підбір пояснювальних ознак здійснимо за методом оптимального вибору предикатів. Його сутність полягає в наступному: спочатку для вихідної сукупності ознак будуються всі одноелементні підмножини, далі всі двоелементні, триелементні і так далі. Остання підмножина буде складатись зі всіх вихідних ознак. Далі для кожної такої підмножини будується показник ємності інформації H_m :

$$H_m = \sum_{i=1}^m h_i,$$
$$h_i = \frac{r_i^2}{1 + \sum_{j \neq i} |r_{ij}|}$$

де r_i – коефіцієнт кореляції між ознакою X_i та пояснювальною ознакою Y ,

r_{ij} – коефіцієнт кореляції між ознакою X_i та іншими ознаками, що входять до складу підмножини ознак, що аналізується;

m – кількість ознак в підмножині, що розглядається.

Найкращою буде та комбінація ознак, для якої значення H_m буде найбільшим.

Для спрощення розрахунків рекомендується обчислити матриці взаємних кореляцій вихідних ознак та вектор кореляцій вихідних ознак і пояснювальної ознаки.

Виконання

Стандартизуємо ознаки.

Стандартизовані ознаки						
U	Z1	Z2	Z3	Z4	Z5	Z6
0,054884	1,568082	1,521489	0,577836	-0,94374	-0,72818	-0,20896
0,137211	1,764338	1,411407	1,266997	-0,42839	-0,06284	-0,09927
1,01536	1,705461	1,678748	1,770615	0,40906	0,898212	-0,01152
1,78374	1,66621	1,584393	1,903146	1,117667	1,119993	0,460153
-1,50932	-1,37575	1,647296	-1,22459	-0,8149	-1,02389	-0,95486
-1,20745	-1,37575	-1,16765	-1,19808	-0,42839	-0,65425	-0,90001
-1,20745	-0,98324	-1,23056	-1,09206	0,602316	0,676431	-0,75742
-0,85071	-0,94399	-0,86886	-1,09206	1,182085	1,267847	-0,6916
-0,63117	-0,37485	-0,72733	-0,0053	-1,52351	-0,80211	-0,47222
-0,24698	-0,37485	-0,69587	0,047711	-1,26583	0,011089	0,218833
-0,38419	-0,4141	-0,72733	0,259761	-1,07257	0,085016	1,140237
-0,38419	-0,45335	-0,79023	0,418798	-0,62164	0,380724	3,355994
-0,16465	-0,45335	-0,3499	-0,98603	0,022547	-1,83708	-0,61482
0,493959	-0,45335	-0,47571	-0,95953	0,731154	-1,3196	-0,52706
0,439074	-0,47298	-0,45998	-0,90651	1,50418	-1,02389	-0,45028
2,607004	-0,4926	-0,49144	-0,82699	2,01953	-0,65425	1,414464
-0,63117	0,272795	-0,11401	0,206748	-0,94374	-0,06284	-0,3735
-0,16465	0,292421	-0,00393	0,286267	-0,42839	0,528577	-0,31865
0,576285	0,469051	0,169054	0,710367	0,215803	1,4157	-0,15412
0,274422	0,4298	0,090424	0,842898	0,666735	1,785335	-0,05539

Запишемо кореляційну матрицю:

Кореляційна матриця							
	U	Z1	Z2	Z3	Z4	Z5	Z6
U	1	0,52363	0,300249	0,435208	0,604957	0,125246	0,360331
Z1	0,52363	1	0,72958	0,89528	0,002908	0,341433	0,107251
Z2	0,300249	0,72958	1	0,609305	-0,04038	0,100578	-0,10721
Z3	0,435208	0,89528	0,609305	1	-0,13818	0,553576	0,325223
Z4	0,604957	0,002908	-0,04038	-0,13818	1	0,194489	-0,00151
Z5	0,125246	0,341433	0,100578	0,553576	0,194489	1	0,183955
Z6	0,360331	0,107251	-0,10721	0,325223	-0,00151	0,183955	1

Підбір пояснювальних ознак здійснимо за методом оптимального вибору предикатів.

Для кожної підмножини обчислюємо H_m

	Hm
1	0,27
2	0,09
3	0,19
4	0,37
5	0,02
6	0,13
7	0,21
8	0,13
9	0,18
10	0,13
11	0,13
12	0,29
13	0,22
14	0,30
15	0,22

X1	X2	X3	X4	X5	X6
0,27	0,09	0,19	0,37	0,02	0,13

X1X2	X3X5
0,16	0,05
0,12	0,01

X1X2X3	X2X3X5
0,08	0,04
0,06	0,04
0,08	0,01

X1X2X3X5	X2X3X4X5
0,06	0,02
0,04	0,00
0,04	0,08
0,16	0,01

X1X2X3X4X5	X2X3X4X5X6
0,06	0,02
0,04	0,09
0,00	0,03
0,07	0,14
0,01	0,05

X1X2X3X4X5X6
0,06
0,02
0,04
0,08
0,00
0,03

Для двох підмножин будуюмо регресії. Методом найменших квадратів обчислюємо параметри, знаходимо емпіричне значення Y і обчислюємо суму квадратів залишків.

Регресийна модель 1				
у	X0	X4	Уемп	Ui^2
185,25	1	45,15	140,78	1977,47
191,1	1	55,47	162,93	793,28
253,5	1	72,24	198,93	2977,41
308,1	1	86,43	229,40	6194,39
74,1	1	47,73	146,32	5215,68
95,55	1	55,47	162,93	4540,72
95,55	1	76,11	207,24	12475,10
120,9	1	87,72	232,16	12379,84
136,5	1	33,54	115,86	426,07
163,8	1	38,7	126,94	1359,01
154,05	1	42,57	135,24	353,71
154,05	1	51,6	154,63	0,33
169,65	1	64,5	182,32	160,51
216,45	1	78,69	212,78	13,47
212,55	1	94,17	246,01	1119,62
366,6	1	104,49	268,16	9689,59
136,5	1	45,15	140,78	18,33
169,65	1	55,47	162,93	45,09
222,3	1	68,37	190,63	1003,19
200,85	1	77,4	210,01	83,93
сума	3627		3627	60826,72

Регресійна модель 2								
у	X0	X2	X3	X4	X5	X6	Уемп	U _i ²
185,25	1	36,72	40,5	45,15	490	22,36	173,514	137,734
191,1	1	35,53	54,54	55,47	580	24,96	223,015	1018,595
253,5	1	38,42	64,8	72,24	710	27,04	265,341	140,221
308,1	1	37,4	67,5	86,43	740	38,22	312,889	22,936
74,1	1	38,08	3,78	47,73	450	4,68	71,915	4,776
95,55	1	7,65	4,32	55,47	500	5,98	106,982	130,681
95,55	1	6,97	6,48	76,11	680	9,36	125,799	914,996
120,9	1	10,88	6,48	87,72	760	10,92	135,076	200,968
136,5	1	12,41	28,62	33,54	480	16,12	122,869	185,796
163,8	1	12,75	29,7	38,7	590	32,5	118,409	2060,321
154,05	1	12,41	34,02	42,57	600	54,34	151,037	9,080
154,05	1	11,73	37,26	51,6	640	106,86	202,645	2361,435
169,65	1	16,49	8,64	64,5	340	12,74	187,250	309,766
216,45	1	15,13	9,18	78,69	410	14,82	212,906	12,562
212,55	1	15,3	10,26	94,17	450	16,64	250,965	1475,705
366,6	1	14,96	11,88	104,49	500	60,84	294,276	5230,833
136,5	1	19,04	32,94	45,15	580	18,46	137,795	1,676
169,65	1	20,23	34,56	55,47	660	19,76	150,205	378,097
222,3	1	22,1	43,2	68,37	780	23,66	180,864	1716,950
200,85	1	21,25	45,9	77,4	830	26	203,249	5,753
сума	3627						3627	16319

Висновок

Отже, при виконанні даної лабораторної роботи було відібрано пояснювальні ознаки і побудовано на основі них економетричні моделі.

Для відбору пояснювальних ознак використаємо метод оптимального вибору предикторів. Аналізуються ознаки за величиною H_m , чим більше її значення, тим кращий відбір.

Після створення і аналізу економетричної моделі виявилось, що найкращою буде комбінація (X2,X3,X4,X5,X6), так як сума квадратів залишків менша при порівнянні регресійних моделей.